Newcastle University

ePrints

Nordahl-Hansen A, Fletcher-Watson S, McConachie H, Kaale A.

Relations between specific and global outcome measures in a social-communication intervention for children with autism spectrum disorder.

***Research in Autism Spectrum Disorders* 2016, 29-30, 19-29.**

DOI link to article:

http://dx.doi.org/10.1016/j.rasd.2016.05.005

Date deposited:

19/05/2016

Embargo release date:

10 December 2017

Relations between specific and global outcome measures in a social-communication intervention for children with autism spectrum disorder

Abstract

Assessment of relevant outcomes is a key challenge in evaluating effects of social-communication interventions. However, few studies have investigated in what ways specific and more global measures may influence reported results of social-communication interventions for children with autism spectrum disorder (ASD). In this study both a specific and a more global autism symptom measure were used to assess effects of a brief social-communication intervention. Fifty-nine children (2-4 years) diagnosed with autistic disorder were assessed with the Joint Engagement (JE) states coding procedure and a preliminary version of the Brief Observation of Social Communication Change (BOSCC). A statistically significant difference was found between intervention and control groups from baseline to intervention endpoint on JE but not on BOSCC. Degree of change on the measures was moderately related, and both were independent of language level and non-verbal mental age. This study adds to the knowledge of what may be expected of different outcome measures and provides suggestions to how measures may be deployed to investigate underlying mechanisms and developmental pathways.

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social communication and interaction as well as restricted and repetitive patterns of behaviours (APA, 2013). The disorder is thought to be caused by complex interactions between genes and environmental factors, but it remains unclear how this interplay influences the behavioural phenotype (Hallmayer et al., 2011; Sandin et al., 2014). Epidemiological studies report worldwide prevalence of ASD to be somewhere within the range of 50 to 70 per 10 000 (Elsabbagh et al., 2012; Fombonne, Quirke, & Hagen, 2011) although some studies report prevalence estimates close to 120 per 10 000 in regions in the UK and the US (Baird et al., 2006; Kogan et al., 2009). Thus, ASD is today recognized as one of the more common developmental disorders.

There has been a substantial increase in ASD related research during the past two decades, including research on early identification and intervention (Charman, 2011; Dawson and Bernier, 2013; Lecavalier, 2016). Knowledge in these areas is important, as an early start for intervention is thought to be crucial for modification of areas of impairments affecting the developmental trajectories of children with ASD (Bradshaw, Steiner, Gengoux, & Koegel, 2015; Webb, Jones, Kelly, & Dawson, 2014). Although results from intervention studies are mixed (Fletcher-Watson, McConnel, Manola, & McConachie, 2014; Howlin, Magiati, Charman, MacLean Jr., 2009; Oono, Honey, & McConachie, 2013), some studies have reported improvement in areas of social-communication functioning (e.g. Green et al., 2010; Kasari, Paparella, Freeman, & Jahromi, 2008; Kaale, Fagerland, Martinsen, & Smith, 2014; Warreyn & Roeyers, 2014; Wetherby et al., 2014).

Interventions aiming to enhance social communication in children with ASD often target a specific set of behaviours or abilities (e.g. joint attention, imitation, shared engagement), but it is implied that gains in specific functioning may lead to improvements in more global autism symptoms (Mundy, Sigman, & Kasari, 1994; Yoder, Bottema-Beutel, Woynaroski, Chandrasekhar, & Sandbank, 2014).

The choice of terminology here is important. Yoder et al. (2014) proposed a matrix where outcome measures were described as ”context-bound” as opposed to ”generalized”, and ”proximal” as opposed to ”distal”, in relation to the intervention tested. The generalized / context-bound dichotomy relates to whether the measurement process was tied to the therapeutic context or not. The proximal / distal dichotomy relates to how far (in an implied linear manner) the evaluated skill lies from the taught skill. Instead, we use the terms ”specific” and ”global”. This is because the transition between the two measurement tools explored in this study represents a broadening of measured traits from a more narrow range directly related to the intervention target, to a wide range representative of the entire constellation of characteristics used in autism diagnosis.

Most studies of the effects of social-communication interventions utilize outcome measures that are specific and related to the intervention targets (Yoder et al, 2014). However, a few high quality social-communication intervention studies have also examined change in global autism symptoms (e.g. Green et al., 2010; Wetherby et al., 2014). The studies indicate different findings derived from specific outcome measures compared to global outcome measures. For instance, Green et al. (2010) conducted a large scale RCT testing the effect of a parent-mediated social-communication intervention, using Autism Diagnostic Observation Schedule-Generic social communication scores (ADOS-G: Lord et al., 2000) as the primary outcome measure. They found that the intervention group improved on symptom scores. However, as improvements were also present in the control group, the treatment effect was statistically non-significant. Nevertheless, significant treatment effects were found on 'blinded' measures more closely related to intervention targets, such as parent interaction style and children's communication initiations to their parent. This is in line with the review by Yoder et al. (2014) where results showed more positive effects of intervention studies using outcome measures specific to intervention targets (e.g. Early Social Communication Scales (ESCS: Mundy et al., 2003)), compared to more global measures (e.g. ADOS: Lord, Rutter, DiLavore, & Risi, 1999).

There is a huge variety of outcome measures employed in autism intervention studies (Bolte & Diehl, 2013; Cunningham, 2012), and these measures differ on multiple aspects, not only whether the measures capture specific or more global aspects of behaviour (Fletcher-Watson & McConachie, 2015; Green et al., 2010). For example, outcome measures may be directly-observed versus informant reports; based on standardised assessment, curriculum-based assessment or free play; amenable to blind rating or not; independent of general developmental level, or not; and designed and proven to be responsive to change over time, or not.

Despite the variety of procedures available, it is clear that to the extent that change in more global outcomes is anticipated, intervention studies should include measures that capture broader aspects of child functioning. This is important to strengthen the validity and importance of conclusions regarding the impact of the intervention on behaviours beyond specific treatment targets

It seems intuitively correct that measures capturing skills and behaviours that are specific to treatment targets may be more sensitive to detection of changes compared to more global measures. However, there is a lack of empirical investigation of this notion, the review by Yoder et al. (2014) being an exception.

In this report, we review two measures closely matched on the criteria listed above (i.e. both directly observed, from a free-play sample, blind rated and designed to capture change over time) which differ in their specificity to the intervention targets.

The specific measure

Measures that focus on change in specific behaviours or skills related to the social-communication domain include, among others, the ESCS (Mundy et al., 2003) and the Communication and Symbolic Behavior Scales–Developmental Profile (CSBS-BP; Wetherby & Prizant, 2002). Although not coding discrete behaviours directly, the Joint Engagement States coding procedure (JE; Bakeman & Adamson, 1984) is another measure that seems to have increasing status as an appropriate outcome measure in evaluation of social-communication interventions for young children with ASD. This coding procedure was developed to measure the quality of the interaction between adult and child through identification of six mutually exclusive engagement states; unengaged, on-looking, person engagement, object engagement, supported joint engagement and coordinated joint engagement (Bakeman & Adamson, 1984). When the child is coded as being in an unengaged state the child seems uninvolved with the adult, toy or activity, although he or she might be scanning the room as to look for something to do. On-looking is a state where the child observes the adult's activity but does not take part. Person engagement is coded when the child is engaging with the adult, but not toys, like in tumble play or face-to face play. In a state of object engagement the child is engaging with a toy but not attending to the adult. Supported joint engagement is coded when the child and adult are actively involved in the same toy, however the child does not show clear signs of acknowledging the adult's presence. Coordinated joint engagement is coded when both child and adult are actively coordinating their attention to a shared toy and to each other. Video recordings of 10-20 minutes of adult-child interaction with a pre-defined set of toys are used to code the various states (see Adamson, Bakeman, & Deckner, 2004; 2012; Bottema-Beutel, Yoder, Hochman, & Watson, 2014 for further developments of the coding procedure). Although the coding procedure was originally developed for typically developing infants and toddlers, it has been used in several longitudinal and intervention studies with preschool-age children with ASD (e.g. Adamson, Bakeman, Deckner, & Romski, 2009; Kasari, Freeman, & Paparella, 2006; Kasari, Gulsrud, Wong, Kwon, & Locke, 2010; Lawton & Kasari, 2012; Lewy & Dawson, 1992; Meek, Robinson, & Jahromi, 2012; Patterson, Elder, Gulsrud, & Kasari, 2014).

The global measure

There are few measures designed to capture changes in global autism symptoms (Anagnostou et al., 2015; McConachie et al., 2015). Due to this shortcoming some researchers have used the ADOS, or an adaptation of the ADOS (e.g. Green et al., 2010; Wetherby et al., 2014) to assess change in autism symptoms following intervention. However, since the ADOS was designed as a diagnostic tool, it may not be reasonable to expect it to sensitively detect differential change between treatment groups, especially not following relatively brief social-communication interventions. To address this problem the Brief Observation of Social Communication Change (BOSCC; Grzadziniski et al., 2016) is being developed as a measure of global change in autism symptoms in young children with ASD. Like the JE coding procedure, the BOSCC is coded based on video observation of natural interaction between adult and child playing with a pre-defined set of toys. BOSCC evolved through expansion of codes from ADOS-2 (Lord et al., 2012) and although more flexible with looser instructions it follows the same scoring logic. The preliminary version of BOSCC used in the present study consists of 16 items related to social communication, play and engagement with objects, stereotypical behaviours, repetitive interests, body mannerisms and self-injurious behaviours.

As the BOSCC and the JE are coded from low-structure dyadic floor play, the naturalistic context also makes them different from, for instance, the CSBS and ESCS that are semi-scripted. Further, although the BOSCC takes frequency of discrete behaviours such as pointing and gaze-alternation into account, the ratings also relies on observations of consistency and quality of the interaction, which is more akin to the JE measure than the CSBS and ESCS. There are also some similarities in the content as coders of both BOSCC and JE need to take into account, for instance, whether the child focuses on, and engages with objects, for how long the child sustains engaged with objects, and whether the child is responding to or acknowledges the adult’s attempts to engage. As well as these conceptual similarities, there are also differences between the two measures in terms of output scale format. The JE-measure codes duration of states whereas the

BOSCC-measure relies on observation of various discrete behaviours that are scored on a six-point-scale

Objective and Hypotheses

This study explores relations between specific and global outcomes in early autism social-communication intervention studies. We characterise outcomes in two categories: 1) measures that are specific to the intervention targets and 2) measures that are global, capturing broader autism symptomatology. Change in specific outcome measures is often proposed to lead, via a developmental cascade effect, to change in global autism symptoms. In order to explore the ways in which specific and global outcome measures may differ in revealing intervention effects, and the relation between these types of measures, we compare two measures by coding the same video recorded parent-child play interactions applying both JE and BOSCC. In order to be useful in measuring outcome across the heterogeneous group of young children with ASD, outcome measures should ideally be independent of general developmental ability and language. Thus, we also examine to what degree the scores are dependent of non-verbal mental age and language level.

Our hypotheses were: (1) treatment effects of a brief social-communication intervention detected by the specific measure (JE) will be greater than those detected by the global measure (BOSCC). This would be in line with earlier studies reporting attenuating effects when measurement content targets broader areas of functioning as compared to specific functioning; (2) children who show change in JE will be more likely to also show change on the BOSCC; (3) change on JE and BOSCC are not explained by general developmental level (indexed by language level and non-verbal mental age).

Method

Design

The present study used baseline and post-intervention data from a previously published randomized controlled trial (RCT) investigating the effects of a brief preschool-based social-communication intervention (Kaale, Smith, & Sponheim, 2012; Kaale et al., 2014). The main focus of the intervention was to increase children's initiations of joint attention and duration of joint engagement using a modification of the intervention manual developed by Kasari et al. (2006). The intervention was delivered by preschool teachers providing two daily 20-minute sessions over eight weeks. Each session included five minutes of tabletop training and 15 minutes of floor play. Tabletop training was preschool teacher led, and mainly focused on creating opportunities for child initiation of joint attention, whereas floor play was child driven and the preschool teachers followed in on the child's activities, trying to facilitate the child to engage jointly with themselves and with toys and objects at hand.

Participants

The original RCT involved 61 children identified by local Child and Adolescent Mental Health Clinics (CAMHCs) in Norway from 2006 to 2008. Two of these children were excluded from the current analysis due to missing information about the variables of interest, leaving a sample size of 59 for the present study. The children met the following inclusion criteria: (i) chronological age of 24–60 months, (ii) ICD-10 diagnosis of childhood autism, and (iii) video recorded parent-child play sample available for re-coding. Exclusion criteria were (i) CNS disorders (e.g. epilepsy, cerebral palsy), and (ii) non-Norwegian speaking parents. All participants were diagnosed with childhood autism by a multi-disciplinary CAMHC team, based on a comprehensive clinical evaluation (interviews and multiple observations by different professionals). Forty-nine children (80%) were tested with Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) and/or Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & LeCouteur, 1994). Missing ADOS/ADI-R was due to site diagnostic practices, not child characteristics. The study participants were functioning within a range of cognitive and language levels (Table 1), and had varied socio-economic backgrounds.

[Insert Table 1 about here]

Measures

Language and developmental level

The children's developmental level was tested with Mullen Scales of Early Learning (MSEL; Mullen, 1997). Overall mental age was calculated based on children's scores on the visual perception, fine motor, and expressive and receptive language subscales, while non-verbal mental age was estimated based on scores on the visual perception and fine motor subscales. The Norwegian standardization of Reynell Developmental Language Scales (RDLS; Hagtvet & Lillestøen, 1985) was used to assess children's language age (receptive and expressive). For children scoring below basal level on the receptive or expressive scales on RDLS, scores on the MSEL language subscales were used to calculate expressive and receptive language age. This approach was justified by high correlation between the attained raw scores on the expressive (*r*= .94) and the receptive (*r*= .97) language scales on the two tests.

Video observations of parent-child play: JE and BOSCC

Video recordings of 10 minutes of parent–child play were used to assess joint engagement and global autism symptoms at baseline and post-intervention through the use of the JE coding procedure, and the BOSCC, respectively. The dyads were given a standard set of toys (a book, two toy phones, a car, blocks, a ball, miniature figures, large marbles, and a stuffed animal) and were instructed to play as they would typically do.

Joint engagement state (JE)

The first author and a research assistant subsequently coded each recorded play session. Both coders were blind to study purpose, group allocation, and testing order. The six engagement states were coded when lasting for at least three seconds and when both the mother and the child were visible on the screen. The joint engagement variable was calculated based on percentage of time in supported and coordinated joint engagement, combined, during the 10 minute recorded play. Collapsing supported and coordinated joint engagement into one joint engagement variable follows the practice of previous studies (e.g. Kasari et al., 2006; Patterson et al., 2014). This was done due to very little time spent in coordinated joint engagement in the dyads.

Inter-rater reliability for the JE was calculated using intra-class correlation coefficients (ICC) based on a random selection of 18% of the mother–child play recordings. ICC was .94 for joint engagement.

Brief Observation of Communication Change (BOSCC)

A preliminary version of the BOSCC was used to code the same baseline and post-intervention video recordings of the mother–child play. The version consists of 16 items, where coding is facilitated by the use of decision trees that leads the coder to assign a score ranging from 0 – 5 on each item. Higher scores indicate greater levels of symptom severity. The coding manual emphasises both frequency and consistency of the observed behaviours for each item.

Thirteen of the 16 items make up the basis for the BOSCC total score. Among these, the first eight items target social behaviours (i.e. the child's ability to gesture, direct vocalizations to others, frequency and quality of responses etc.). These can also be summed to provide a 'BOSCC Social' sub-score. Items 9-13 target stereotypical and repetitive behaviours, body mannerisms, unusual sensory interests and the child's engagement with materials and play with objects. In addition to the 13 items making up the total score, three items (14-16) assess the child's activity level, disruptive/irritable behaviour and anxious behaviour but they are not included in the total score.

The first author was trained on the BOSCC, and on-going reliability was monitored, using a separate data set. This was a sample of parent-play recordings collected as part of a distinct, UK-based RCT of a therapeutic iPad app (Fletcher-

Watson et al., 2015). The UK RCT sample consisted of 61 children with childhood autism aged less than 6 years old at baseline. Parent-child play recordings were collected at baseline and post-intervention (excluding dropouts etc.) resulting in a library of approximately 115 recordings available for use in training. The Norwegian and UK intervention participants were comparable in terms of age, ADOS score, mental age and average BOSCC scores at baseline. We randomly selected parent-child play clips from the UK RCT library using eleven of these for training (A.N-H and S.F-W) and to achieve the required level of agreement for use of the BOSCC, which states *"coders must be within one point for ≥80% of items … and total change scores … must be within three points. This requirement must be met for three consecutive videos".* The BOSCC manual further recommends that *"double coding of videos is completed at least every fifth video to ensure inter-rater agreement across time".* Therefore a further 23 recordings were used to evaluate on-going inter-rater agreement throughout the period that the primary data set were being coded. Inter-rater reliability for the BOSCC was calculated using ICC, which was .99 for BOSCC total and .99 for BOSCC social.

The inter-rater reliability reported here is similar to that found in two other studies using the preliminary version of the BOSCC (Fletcher-Watson et al., 2015; Kitzerow, Teufel, Wilker, & Freitag, 2015), and very similar to the inter-rater reliability reported for the final version of the BOSCC (ICC= 0.97 to 0.98; Grzadzinski et al., 2016). High internal consistency has been found for the BOSCC social subscale (Cronbach's α= 0.83), but lower internal consistency was found for the items measuring restricted and repetitive behaviours (Cronbach's α= 0.41). Additionally, Grzadzinski and colleagues report that the BOSCC has high test-retest reliability and indications of convergent validity with other measures assessing social communication (e.g. Vineland Adaptive Behavior Scales; VABS; Sparrow et al., 2005).

Assessment Procedure

The measures used in the present study were collected as part of a more comprehensive baseline and post-intervention assessment in the original intervention study. Baseline assessments of language and developmental level and video recording of mother-child play were done during one day at the local CAMHC by a tester independent of the research group and blind to the children's group allocation. Prior to the assessment, parents completed a questionnaire about demographic information. The video recording of mother-child play was repeated at post-intervention in the children's preschools. The change of assessment setting was done to limit travelling burden for the participating families. The study was approved by The Norwegian National Committee for Research Ethics. Written consent was obtained from parents and preschools.

Statistical analyses

Differences between the intervention and the control groups on BOSCC social, BOSCC total and JE from baseline to post-intervention were estimated using independent sample t-tests with effect size estimates based on Cohen's *d*. Next, we used the reliable change index (RCI: Jacobson & Truax, 1991) to categorize each child as having achieved, or not achieved, a reliable change on BOSCC total and JE. The RCI is a measure of clinically significant change, which indicates whether an individual's change score exceeds what would be expected on the basis of normal variability. In other words, a certain amount of variability can be expected when measuring the same quantity at two different time points. The RCI determines whether a child’s change is greater than this background variability. Other variants of the RCI are also available when test variables are more susceptible to practice effects as is often the case for standardized neuropsychological tests (Busch, Lineweaver, Ferguson, & Haut, 2015; Chelune, naugle, Lüders, Sedlak, & Awad; Parsons, Notebaert, Shields, & Guskiewicz, 2009). Each child’s change score from baseline to post-intervention was referenced against the overall correlation between baseline and post-intervention scores for the whole control group. McNemar’s test of correlated proportions was then used to assess the difference in probability of obtaining a reliable change score on BOSCC compared to JE for control and intervention group separately. A significant result on this test would indicate that the probability of a child exhibiting a clinically significant change on one measure was greater than for the other measure.

Each child’s change score for BOSCC total and JE were mapped graphically to illustrate individual variation and the relationship between changes on the two measures. As the BOSCC total and JE are scored on different scales, adjusted *Z*-scores for each measure are presented. Pearson’s correlation was used to assess the association between JE, BOSCC total and BOSCC social change scores, and whether changes on these measures were independent of children’s language age and non-verbal mental age at baseline. The statistical analyses were done using SPSS 22.0 except for McNemar’s test where an online calculator was used (Vassarstats.net). All reported *p*-values are two-sided.

Results:

BOSCC and JE change scores: group level

The children in the intervention group showed a mean JE change of 13.8% compared to -1.3% for the control group. This group difference was statistically significant (p= 0.013, d= 0.67) (Table 2). In contrast, no treatment effect was identified for either BOSCC total or BOSCC social (respectively p= 0.244, d= 0.31 and p= 0.196, d= 0.34).

[Insert Table 2 about here]

Distribution of reliable change

Table 3 illustrates the numbers of children in the intervention group showing reliable change according to one measure, neither of the measures, or both measures. The McNemar's test for correlated proportions was statistically significant (p=< 0.01)[1] indicating that within the intervention group the proportion of children with a reliable change score was different between the two measures. This reiterates the finding above that when assessed with JE, being specific and directly related to the intervention targets, there was a higher likelihood of attaining a reliable change than when assessed with the global measure, the BOSCC.

[Insert table 3 about here]

[1] It is not possible to calculate a confidence interval for this result as the cell count in one cell is zero

Table 4 illustrates the numbers of children in the control group showing reliable change according to one measure, neither of the measures, or both measures. Analysing these results with McNemar's test for correlated proportions, no statistically significant difference was found (p= 0.51, 95% CI = 0.33, 4.66). In other words, the proportions of children showing reliable change on the measures are not consistently distributed.

[Insert table 4 about here]

BOSCC and JE change scores: individual scores

To further illustrate how change on JE may relate to change on the BOSCC, Figures 1 and 2 shows individual change scores on both measures in the intervention and the control group, respectively. We can see in Figure 1 that among the 20 intervention group children showing reliable change on JE only four showed reliable change on BOSCC, and these four children did not show the largest JE increase.

Inspecting the control group data in Figure 2 we can see that there appear to be more children showing a positive (although not reaching RCI cut-off) change on BOSCC compared to JE.

[Figures 1 and 2 in about here]

Correlations between change scores on JE and BOSCC

Moderate but statistically significant correlations were identified between the change from baseline to post-intervention on the two measures (JE and BOSCC total: Pearson's r = .37, p= .001; JE and BOSCC social: Pearson's r = .28, p= .05).

Correlations with non-verbal mental age and language age.

There were close to zero correlations between children's baseline non-verbal mental age and language and BOSCC total, BOSCC social and JE change scores (Table 5).

[Table 5 in about here]

Discussion

The present study aimed to investigate relations between specific and global outcomes in early autism intervention studies. This was done by comparing results from the JE coding procedure and a preliminary version of the BOSCC which were coded from video-recorded parent-child play at baseline and post-intervention in a social-communication intervention trial.

Overall, the joint engagement measure showed a significant difference in change scores between experimental groups. In contrast, when the BOSCC measure was applied to the same recordings, no difference between groups was identified. This was in accordance with our initial hypothesis suggesting that the global measure would not detect intervention effects to the same extent as the specific measure. In further support of the hypothesis, the results from the RCI-analyses, particularly for the intervention group, showed a significantly higher likelihood of reaching a reliable change cut-off score when scored with the specific measure. Thus, this study coincides with other authors' interpretations (Green et al., 2010; Yoder et al., 2014) that outcome measures specific to the intervention target have a higher likelihood of detecting intervention effects compared to global outcome measures. It is important to emphasise that the post-intervention assessment was conducted only three months after baseline assessment. Treatment effects on global autism symptom severity measured by BOSCC might be more apparent at a later time point, which would be compatible with a model of intervention that emphasised cascading effects.

Cascading intervention effects

Often social-communication interventions for children with ASD build on theoretical accounts of developmental cascades (see e.g. Masten and Cicchetti, 2010; Sameroff, 2000) where intervention effects first deliver improvements on behaviours or skills specific to the intervention targets and then these are translated into more global changes in autism symptom severity. Thus, we hypothesized that a positive reliable change in JE (an increase of joint engagement) would be a pre-requisite for a positive reliable change on the BOSCC (a decrease in autism symptoms). No clear patterns in the results emerged to support a developmental cascade from JE to broader autism symptoms in the time-frame of the present study. However, studies using the same intervention content has reported effects on more distal measures of language on long-term follow-up (Kasari, Gulsrud, Freeman, Paparella, & Hellemann, 2012; Kasari, Paparella, Freeman, & Jahromi, 2008). This highlights the need for follow-up studies, as behaviours and skills taught in interventions are likely to take time to be manifested in broader abilities.

Revealing mechanisms of change using multiple outcome measures

In line with the intention of the BOSCC developers, and critical for outcome measurement in heterogeneous groups of young children with ASD, we found that BOSCC change scores from baseline to post-intervention seems to be independent of children's initial language level and non-verbal mental age. This was also the case for the JE measure. Thus, our data suggest that changes on the JE and BOSCC measures are not attributable to underpinning developmental abilities.

There was a positive relation between the two measures as indicated by moderate correlations between change scores. This could indicate that the two measures are, to some degree, tapping in to related phenomena. However visualisations of individual data show that this relationship is far from direct, since, for example, children showing the largest improvements on the BOSCC did not also show the greatest increase in joint engagement. Thus, while it may be that joint engagement improvement can lead to improvement in global autism symptoms, there are also other factors mediating this relationship. Sophisticated research designs are required, using both specific measures as well as more global measures recorded at appropriate time points, to give opportunities for identifying underlying mechanisms of change (Vivanti, Prior, Williams, & Dissanayake, 2014).

Positive global change in both groups

The change score graphs displaying individual change for each measure illustrate that almost all children showed a positive BOSCC change score regardless of group allocation, indicating a decrease in severity of autism symptoms from baseline to post-intervention. This was not the case for JE change scores, which were clear in the intervention group and close to zero in the control group. The presence of positive change in ratings of autism symptoms also in control group participants (Estes et al., 2015; Dawson et al., 2010; Green et al., 2010; Wetherby et al., 2014) presents a significant challenge to the pursuit of statistically significant intervention effects and the design of a measure that is sensitive to treatment effects.

One approach to the problem of non-significant treatment effects in broader autism symptoms is long-term follow-up studies and an emphasis on clinical significance and reliable change in addition to statistical significance. Adding these dimensions can provide a more general understanding of the impact of intervention trial findings (Cicchetti et al., 2011) as well as providing routes to more meaningful analyses at the individual level.

Strengths and limitations

A strength of this study is the use of data from an RCT with a fairly large sample size. Both measures were coded using the same video recordings, decreasing different sources of error such as fluctuation of motivation. That it is possible to code the same video observation with measurement techniques targeting both specific and global variables is potentially time- and cost-effective, as well as reducing the burden of testing for participants. Also, the use of video-based assessments of natural interaction increases ecological validity, which can be a problem in highly standardized measurement procedures (Bacon et al., 2014).

A different, but highly similar sample of children with ASD was used to assess inter-rater reliability for the BOSCC. This could be a confounding factor that contributes to the differences found between the two measures. Nevertheless, the inter-rater reliability for both measures was very high, and the inter-rater reliability estimated for the BOSCC in this study matches the estimates from the two previously published studies (Grzadzinski et al., 2016; Kitzerow et al., 2015). Thus, it seems that the JE coding and the BOSCC measure are both viable candidates for use in intervention evaluation research for children with ASD. Although the BOSCC measure used in this study was a preliminary version and has undergone changes by its developers (Grzadzinski et al., 2016) the main aim of this study was to use the BOSCC as an exemplar of a measure of global outcome that could be compared to a specific outcome measure. Also, elaborations of, and additions to, coding procedures within the JE-paradigm continue to be proposed in other studies (e.g. Adamson et al., 2004; Adamson, Bakeman, Deckner, & Nelson, 2012; Bottema-Beutel et al., 2014; Patterson et al., 2014).

Future directions

In line with the conclusions of Yoder et al. (2014), this study shows that there is a greater likelihood of finding effects of a social-communication intervention when using a specific compared to a global outcome measure. The use of specific outcomes that in part may overlap with the intervention focus is by no means bad practice (Gersten et al., 2005). The measure selected will depend on the study hypotheses and intervention targets. However, a common hope of parents and providers of social-communication interventions for children with ASD is that enhancement of specific areas of functioning may lead to a decrease in more global autism symptoms. To enhance the validity of conclusions, it is necessary that intervention studies use both specific and global measures of autism symptoms. Further, to investigate which mechanisms contribute to developmental cascades, sophisticated statistical analyses such as mediation analyses should be considered (see for instance Pickles et al., 2015). Longitudinal intervention studies incorporating both specific and global measures could enhance the limited knowledge about underlying mechanisms of change (Lecavalier, 2016), highlighting how behaviours and skills taught in interventions may alter downstream development at a more global level.

Acknowledgements:

We sincerely thank Rebecca Grzadzinski and Professor Cathy Lord, and the Western Psychological Services for permission to use the prepublication of the BOSCC for this study. We also thank the COST ESSEA (Action BM1004) for funding.

References

Adamson, L.B., Bakeman, R., & Deckner, D.F. (2004). The development of symbol-infused joint engagement. *Child Development* 75(4): 1171-1187.

Adamson, L.B., Bakeman, R., Deckner, D.F., & Nelson, P.B. (2012). Rating parent-child interactions: joint engagement, communication dynamics, and shared topics on autism, down syndrome, and typical development. *Journal of Autism and Developmental* Disorders 42(12): 2622-2635.

Adamson, L.B., Bakeman, R., Deckner, D.F., & Romski, M. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism and Developmental Disorders* 39(1): 84-96.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (fifth ed.).* Arlington, VA: American Psychiatric Publishing.

Anagnostou, E., Jones, N., Huerta, M., Halladay, A.K., Wang, P., Scahill, L., … Dawson, G. (2015). Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism* 19(5): 622-636.

Bacon, E.C., Dufek, S., Schreibman, L., Stahmer, A.C., Pierce, K., & Courchesne, E. (2014). Measuring outcome in an early intervention program for toddlers with autism spectrum disorder: use of curriculum-based assessment. *Autism Research and Treatment,*1-9, http://dx.doi.org/10.1155/2014/964704

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the special needs and autism project (SNAP). *The Lancet,* 368, 210-215.

Bakeman, R., & Adamson, L.B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development* 55(4): 1278-1289.

Bolte, E.E., & Diehl, J.J. (2013). Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. *Journal of autism and developmental disorders*, *43*(11), 2491-2501.

Bottema-Beutel, K., Yoder, P.J., Hochman, J.M., & Watson, L.R. (2014). The role of supported joint engagement and parent utterances in language and social communication development in children with autism spectrum disorder. *Journal of Autism and developmental Disorders, 44,* 2162-2174.

Bradshaw, J., Steiner, A.M., Gengoux, G., & Koegel, L.K. (2015). Feasibility and effectiveness of very early intervention for infants at-risk for autism spectrum disorder: A systematic review. *Journal of Autism and Developmental Disorders*, *45*, 778-794.

Busch, R.M., Lineweaver, T.T., Ferguson, L., & Haut, J.S., (2015). Reliable change indices and standardized regression-based change score norms for evaluating neuropsychological change in children with epilepsy. *Epilepsy & Behavior, 47,* 45-54.

Charman, T. (2011). Commentary: Glass half full or half empty? Testing social communication interventions for young children with autism – reflections on Landa, Holman, O’neill, and Stuart (2011). *Journal of Child Psychology and Psychiatry*, *52*, 22-23.

Chelune, G.J., Naugle, R.I., Lüders, H., Sedlak, J., & Awad, I.A., (1993). Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology, 7,* 41-52.

Cicchetti, D.V., Koenig, K., Klin, A., Volkmar, F.R., Paul, R., & Sparrow, S. (2011). From Bayes through marginal utility to effect sizes: a guide to understanding the clinical and statistical significance of the results of autism research finding. *Journal of Autism and Developmental Disorders, 41*, 168-174.

Cunningham, A.B., (2012). Measuring change in social interaction skills of young children with autism. *Journal of Autism and Developmental Disorders, 42,* 593-605.

Dawson, G., & Bernier, R. (2013). A quarter century of progress on the early detection and treatment of autism spectrum disorder. *Development and Psychopathology, 25,* 1455-1472.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J. … Varley, J. 2010. Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics, 125,* 17-23.

Elsabbagh, M., Divan, G., Koh, J., Kim, J.S., Kauchali, S., Marcin, C., … Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research, 5,* 160-179.

Estes, A., Munson, J., Rogers, S.J., Greenson, J., Winter, J., & Dawson, G. (2015) Outcomes of early intervention in 6-year-old children with autism spectrum disorder. *Journal of the Academy of Child & Adolescent Psychiatry,* doi: 10.1016/j.jaac.2015.04.005

Fletcher-Watson, S., and McConachie, H. (2015). The search for an early intervention outcome measurement tool in autism. *Focus on Autism and Other Developmental Disabilities*, doi: 10.1177/1088357615583468

Fletcher-Watson, S., McConnell, F., Manola, E., & McConachie, H. (2014). Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD). *The Cochrane Library*.

Fletcher-Watson, S., Petrou, A., Scott-Barrett J., Dicks, P., Graham, C., O'Hare, A. … McConachie, H. (2015) A trial of an iPad^{TM} intervention targeting social communication skills in children with autism. *Autism*, DOI: 10.1177/1362361315605624.

Fombonne, E., Quirke, S., & Hagen, A. (2011). Epidemiology of pervasive developmental disorders. In Amaral D.G., Dawson, G. G. DH (Eds.), *Autism spectrum disorders* (pp. 90-111). Oxford: Oxford University Press.

Gersten, R., Fuchs, L., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M.S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*, 149-164.

Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P. … the PACT consortium. (2010). Parent-mediated communication focused treatment in children with autism (PACT): A controlled randomised controlled trial. *The Lancet*, *375*, 2152-2160.

Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A., & Lord, C. (2016). Measuring changes in social communication behaviors: Preliminary development of the Brief Observation of Social Communication Change (BOSCC). *Journal of Autism and Developmental Disorders*, doi:10.

1007/s10803-016-2782-9

Hagtvet, B. & Lillestølen, R., (1985). *Reynell Språktest.* Oslo: Universitetsforlaget.

Hallmayer, Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T. … Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry, 68,* 1095-1102.

Howlin P, Magiati I, Charman T., & MacLean Jr. W.E. (2009) Systematic review of early intensive behavioral interventions for children with autism. *American Association on Intellectual and Developmental Disabilities* 114(1): 23-41.

Jacobson, N.S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-9.

Kaale, A., Fagerland, M.W., Martinsen, E.W., & Smith, L. (2014). Preschool-based social communication treatment for children with autism: 12-month follow-up of a randomized trial. *Journal of the American Academy for Child and Adolescent Psychiatry, 53*, 188-198.

Kaale, A., Smith, L., & Sponheim, E. (2012). A randomized controlled trial of preschool-based joint attention intervention for children with autism. *Journal of Child Psychology and Pscychiatry, 53,* 97-105.

Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: a randomized controlled intervention study. *Journal of Child Psychology and Psychiatry, 47*, 611-620.

Kasari, C., Gulsrud, A., Freeman, S., Paparella, T, Hellemann, (2012). Longitudinal follow up of children with autism receiving targeted interventions on joint attention and play. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*, 487-495.

Kasari, C., Gulsrud, A.C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders, 40,* 1045-1056.

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L.B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology, 76,* 125-137.

Kitzerow, J., Teufel, K., Wilker, C., & Freitag, C.M. (2015). Using the brief observation of social communication change (BOSCC) to measure autism-specific development. Autism Research, DOI:10.1002/aur.1588.

Kogan, M.D., Blumberg, S.J., Schieve L.A., Boyle, C.A., Perrin, J.M., Ghandour, R.M., ... van Dyck, P.C. (2009) Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. Pedatrics, 124, 1395-1403.

Lawton, K., & Kasari, C. (2012). Teacher-implemented joint attention intervention: Pilot randomized controlled study for pre-schoolers with autism. *Journal of Consulting and Clinical Psychology, 80,* 687-693.

Lecavalier, L. (2016). Autism spectrum disorder clinical trials: One step at a time. *Autism, 20,* 131-133.

Lewy, A.L., & Dawson, G. (1992). Social stimulation and joint attention in young autistic children. *Journal of Abnormal Child Psychology, 20,* 555-566.

Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30,* 205-223.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999) Manual: Autism diagnostic observation schedule. *Los Angeles, CA: Western Psychological Services.*

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2).* Torrance: Western Psychological Services.

Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24,* 659-685.

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., … Yoder, P. (2005). Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 35,* 695–708.

Magiati, I., Moss, J., Charman, T., & Howlin, P. (2011). Patterns of change in children with autism spectrum disorders who received community based comprehensive interventions in their pre-school years: A seven year follow-up study. *Research in Autism Spectrum Disorders, 5,* 1016-1027.

Masten, A.S., and Cicchetti D. (2010) developmental cascades. *Development and Psychopathology, 22,* 491-495.

McConachie, H., Parr, J.R., Glod, M., Hanratty, J., Livingstone, N., Oono, I.P., … Williams, K. (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder. *Health Technology Assessment, 19,* DOI: 10.3310/hta19410.

Meek, S.E., Robinson, L.T., & Jahromi, L.B. (2012). Parent-child predictors of social competence with peers in children with and without autism. *Research in Autism Spectrum Disorders, 6,* 815-823.

Mullen, E. M., (1997). Mullen scales of early learning. Circle Pines, MN: American Guidance Service.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *A manual for the abridged early social communication scale (ESCS): Preliminary manual.* Retrieved from http://www.ucdmc.ucdavis.edu/mindinstitute/ourteam/faculty_staff/escs.pdf

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology, 6,* 389-401.

Oono, I.P., Honey, E.J. & McConachie, H. (2013). Parent-mediated early intervention for young children with autism spectrum disorders (ASD). *Evidence-Based Child Health: A Cochrane Review Journal, 8,* 2380–2479.

Parsons, T.D., Notebaert, A.J., shields, E.W., & Guskiewicz, K.M. (2009). Application of reliable change indices to computerized neuropsychological measures of concussion. *International Journal of Neuroscience, 119,* 492-507.

Patterson, S.Y., Elder, L., Gulsrud, A., & Kasari, C. (2014). The association between parental interaction style and children's joint engagement in families with toddlers with autism. *Autism, 18,* 511-518.

Pickles, A., Harris, V., Green, J., Aldred, C., McConachie, H, Slonims, V., ... The PACT Consortium. (2015). Treatment mechanisms in the MRC preschool autism communication trial: implications for study design and parent-focussed therapy for children. *Journal of Child Psychology and Psychiatry, 56,* 162-170.

Sameroff, A.J. (2000). Developmental systems and psychopathology. *Development and Psychopathology, 3,* 297-312.

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larson, H., Hultman, C.M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA Psychiatry, 311,* 1770-1777.

Smith, T., Scahill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., Rogers, S., & Wagner, A. (2007). Designing research studies on psychosocial interventions in autism. *Journal of Autism and Developmental Disorders, 37,* 354-366.

Vivanti, G., Prior, M., Williams, K., & Dissanayake, C. (2014). Predictors of outcome in autism early intervention: why don't we know more? *Frontiers in Pediatrics,* 1-10 doi,10.3389/fped.2014.00058

Warreyn, P., & Roeyers, H. (2014). See what I see, do as I do: Promoting joint attention and imitation in pre-schoolers with autism spectrum disorder. *Autism, 18,* 658-671.

Webb, S.J., Jones, E.J.H., Kelly, J., & Dawson, G. (2014). The motivation for very early intervention for infants at high risk for autism spectrum disorders. *International Journal of Speech-Language Pathology, 16,* 36-42.

Wetherby, A.M. and Prizant, B.M. (2002). *Communication and Symbolic Behavior Scales: Developmental Profile, 1st normed ed.* Baltimore MD, US: Paul H Brookes Publishing.

Wetherby, A.M., Guthrie, W., Woods, J., Schatschneider, C., Holland, R.D., Morgan, L., & Lord, C. (2014). Parent-implemented social intervention for toddlers with autism: An RCT. *Pediatrics, 134,* 1084-1093.

Yoder, P.J., Bottema-Beutel, K., Woynaroski, T., Chandrasekhar, R., & Sandbank, M. (2014). Social communication intervention effects vary by dependent variable type in pre-schoolers with autism spectrum disorders. *Evidence-Based Communication Assessment and Intervention, 7,* 150-174.

Figure 1 change score graph intervention group

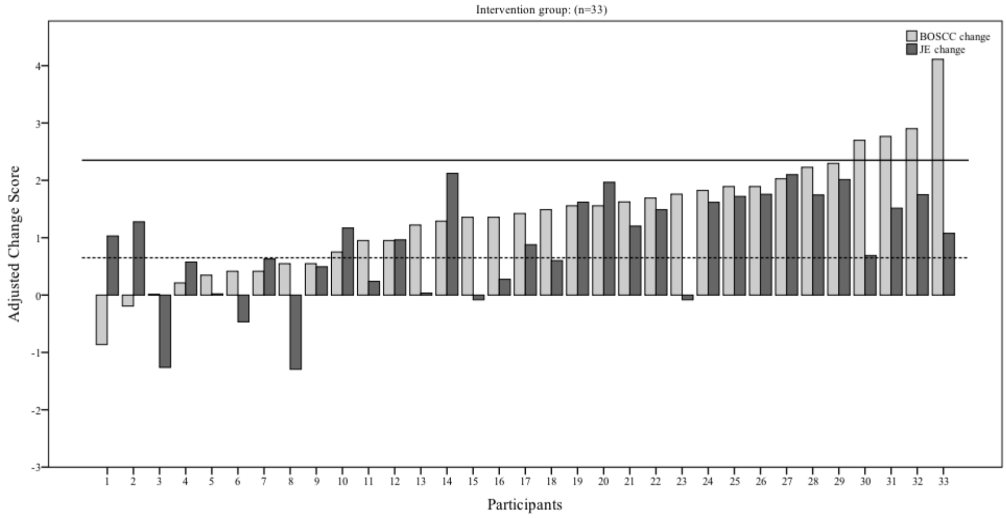

Figure 1. Individual data showing degree of change between baseline and post-intervention in BOSCC and JE. Dotted line is RCI cut-off for JE. Straight line is RCI cut-off for BOSCC. BOSCC = Brief Observation of Social Communication Change; JE = Joint Engagement coding procedure; RCI = Reliable Change Index

Figure 2 change score graph control group

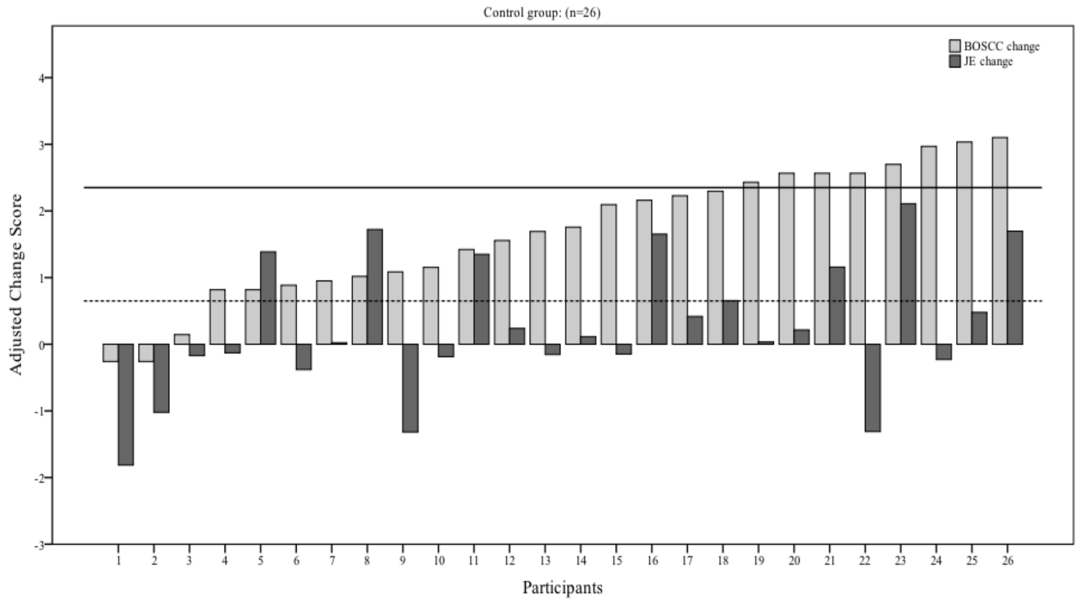

Figure 2. Individual data showing degree of change between baseline and post-intervention in BOSCC and JE. Dotted line is RCI cut-off for JE. Straight line is RCI cut-off for BOSCC. BOSCC = Brief Observation of Social Communication Change; JE = Joint Engagement coding procedure; RCI = Reliable Change Index

Table 1: Child characteristics for experimental groups at baseline (n=59)

	Intervention group (n=33)	Control group (n=26)
	Mean (SD)	Mean (SD)
Age (months)	47.5 (8.4)	50.0 (8.4)
Mental age, months[1]	26.2 (11.6)	29.2 (13.4)
Language age, months[2]	20.0 (10.1)	25.9 (11.7)
Gender male n (%)	25 (75.8%)	22 (84.6%)

[1]Mullen Scales of Early Learning (MSEL)

[2]Receptive and expressive language primarily based on Reynell Developmental Language Scales (RDLS), but for scores < 4 on the stanine scale for 1.5 yrs on RDLS, language age was based on MSEL.

Table 2 Intervention effects

Table 2: Effects of intervention for intervention- and control –group as measured by change scores

	Intervention group [n=33]			Control group [n=26]				
	Baseline mean (*SD*)	Post mean (*SD*)	Change mean *(SD)*	Baseline mean (*SD*)	Post mean (*SD*)	Change mean *(SD)*	Effect size estimate	*t*-test *p*-value
Joint engagement (JE)	45.1 *(23.4)*	58.9 *(21.2)*	13.8 *(21.4)*	50.2 *(21.7)*	48.9 *(20.2)*	-1.3 *(23.8)*	0.67	0.013
BOSCC total (1-13)	36.1 *(13.4)*	31.5 *(13.3)*	-4.6 *(7.5)*	32.1 *(11.8)*	25.2 *(11.8)*	-6.9 *(7.3)*	0.31	0.244
BOSCC social (1-8)	27.3 *(10.4)*	24.9 *(10.8)*	-2.4 *(6.6)*	25.1 *(9.2)*	20.4 *(10.1)*	-4.7 *(6.8)*	0.34	0.196

NB: Minus change (-) on BOSCC scores and plus (+) on JE are both gains.

Table 3 2x2 table intervention

Table 3. Intervention group reliable change BOSCC and JE

			JE–Reliable Change index		
			No Reliable change	Reliable change	Total
BOSCC-Reliable Change index	No reliable change	Count	13	16	29
	Reliable change	Count	0	4	4
		Total	13	20	N=33

Table 4 2x2 table control group

Table 4. Control group reliable change BOSCC and JE

			JE–Reliable Change index		
			No Reliable change	Reliable change	Total
BOSCC-Reliable Change index	No reliable change	Count	14	4	18
	Reliable change	Count	5	3	8
		Total	19	7	N=26

Table 5. BOSCC change scores (total and social) and JE change scores correlations with non-verbal mental age and language age (n=59)

	JE change scores	BOSCC Total change	BOSCC Social Change
Non-verbal mental age	-.07	-.01	-.07
Language Age	-.05	.06	.16

Pearson's *r* correlations